ERJ open research
THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Early View

Original research article

An updated approach to determine minimal clinically important differences in idiopathic pulmonary fibrosis

Mohleen Kang, Srihari Veeraraghavan, Greg S. Martin, Jordan A. Kempker

Please cite this article as: Kang M, Veeraraghavan S, Martin GS, *et al*. An updated approach to determine minimal clinically important differences in idiopathic pulmonary fibrosis. *ERJ Open Res* 2021; in press (https://doi.org/10.1183/23120541.00142-2021).

This manuscript has recently been accepted for publication in the *ERJ Open Research*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJOR online.

Copyright ©The authors 2021. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Title: An Updated Approach to Determine Minimal Clinically Important Differences In Idiopathic Pulmonary Fibrosis

Authors:
Mohleen Kang, MD[1]; Srihari Veeraraghavan, MD[1]; Greg S Martin, MD, MSc[1]; Jordan A Kempker, MD, MSc[1]

1. Emory University School of Medicine
   Division of Pulmonary, Allergy, Critical Care and Sleep Medicine
   615 Michael St. NE Ste 205, Atlanta, GA 30322
   United States

Corresponding Author:
Dr. Mohleen Kang, MD
615 Michael St. NE Ste 205, Atlanta, GA 30322, United States
Phone: 404-712-8286
Email: mkang30@emory.edu

Other Authors' Emails:
Dr. Srihari Veeraraghavan, MD
Email: veeraraghavan@emory.edu

Dr. Greg S Martin, MD, MSc
Email: greg.martin@emory.edu

Dr. Jordan A Kempker, MD, MSc
Email: jkempke@emory.edu

Declaration of Interest: Dr. Kang was supported by her Division's T-32 grant 5T32HL116271-07 (PI: Dr. David Guidot) during this study. Dr. Veeraraghavan was received personal fees from Boehringer Ingelheim for serving on their advisory board and has also received research grant support from Fibrogen, Bellerophon, Biogen, Nitto Denko, Pliant, Galapagos, and Galecto. Dr. Martin received support from the National Institutes of Health through the National Center for Advancing Translational Sciences UL1TR002378, National Institute of Biomedical Imaging and Bioengineering U54 EB-027690, and the Office of the Director OT2 OD-026551; as well as serving as a consultant to Genentech and Grifols. Dr. Kempker received support from the Agency for Healthcare Quality and Research under Award Number K08HS025240 and has worked as a consultant for Grifols. Inc.

Take Home Point: Current consensus approaches recommend anchor-based estimation of MCID over distribution-based methods. MCID values of 6MWD, SGRQ, SF-36, UCSD SOBQ using only anchor-based method were higher than previously reported values.

Abstract:

Introduction: Current medications for idiopathic pulmonary fibrosis (IPF) have not been shown to have impact on patient related outcome measures (PROMs) highlighting the need for accurate Minimal Clinically Important Differences (MCID) values. Recently published consensus standards for MCID studies support using anchor-based over distribution-based methods. The aim of this study was to estimate MCID values for worsening in IPF using only an anchor-based approach.
Methods: We conducted secondary analyses of three randomized controlled trials with different inclusion criteria and follow-up intervals. The Health Transition question in the Short Form Health Survey 36 (SF-36) questionnaire was used as the anchor. We used receiver operating curve to assess responsiveness between the anchor and ten variables (four physiologic measures and six PROMs). We used an anchor-based method to determine the MCID values of variables that met the responsiveness criteria (area under the curve ≥ 0.70).
Results: Six minute walk distance (6MWD), the St. George's Respiratory Questionnaire (SGRQ), physical component score of SF-36 (SF-36 PCS), and University of California, San Diego, Shortness of Breath Questionnaire (UCSD SOBQ) met the responsiveness criteria. The MCID value for 6MWD was -75 meters. The MCID value for SF-36 PCS was -7 points. MCID value for SGRQ was 11points. MCID value for the UCSD SOBQ was 11 points.
Conclusions: The MCID estimates of 6MWD, SGRQ, SF-36, UCSD SOBQ using only anchor-based methods were considerably higher compared to previously proposed values. A single MCID value may not be applicable across all classes of disease severity or durations of follow-up time.

Keywords: Idiopathic pulmonary fibrosis; Interstitial Lung disease; Minimal clinically important difference; Patient centered research.

Introduction:

Idiopathic Pulmonary Fibrosis (IPF) is a chronic fibrosing lung disease that is progressive and has a median survival of 2-3 years after diagnosis [1]. The disease progression is associated with increased symptom burden and is punctuated by episodic acute exacerbations that can lead to hospitalization and acute respiratory failure. Mortality and hospitalization are meaningful but challenging primary endpoints in IPF as they require large sample sizes and long follow-up periods [2]. Therefore, measures of lung function such as Forced Vital Capacity (FVC) are more feasible endpoints for drug trials. There are currently two pharmacologic treatment options, Pirfenidone and Nintedanib, which have been shown to decrease the rate of annual decline of FVC [3-6]. Neither of these medications, however, has shown an impact on patient reported outcomes measures (PROMs) as measured by the St. George's Respiratory Questionnaire (SGRQ) or the University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ). This raises an important issue as to what minimal clinically important difference (MCID) in outcome measures such as FVC and SGRQ would be associated with clinically meaningful change in patients.

MCID is a threshold value for a change in a measure considered meaningful by the patient and which, per Jaeschke who first defined the concept in 1989, "would mandate, in the absence of troublesome side-effects and excessive cost, a change in patient management''[7]. MCID is often used in trial design to estimate effect size for sample size calculation and in evaluating the clinical importance of trial results. For instance, a statistically significant difference in a primary endpoint such as FVC between the treatment and control groups may not be clinically important for patients if it falls below the MCID value of that primary endpoint. MCID values have traditionally been determined by three different methods: anchor-based, distribution-based, and expert opinion. The anchor-based methods estimate MCID as the quantity of change in a measure that is associated with patient's report of minimal improvement or worsening i.e., the anchor. Distribution-based methods use statistical methods to determine the minimal change that can be detected beyond statistical error without incorporating patient input. Expert opinion incorporates formal or informal clinician judgments as the MCID value. While there is no gold standard methodology to determine MCID values, proposed tools and consensus approaches support anchor-based over distribution-based methods [20, 21].

Among the 10 articles that have studied MCID values of various measures in IPF, there are some limitations [13-22]. Nine out of the ten studies utilized distribution-based methods to calculate MCID [13-18, 20-22]. Distribution-based methods do not incorporate patient input and, therefore, may not necessarily reflect patient-centered differences [23, 24]. Additionally, while these studies also used anchor-based methods, some of the studies used mortality and or hospitalization as anchors, which while clinically important to patients, may determine "maximal" rather than "minimal" important changes [15-17]. Similarly, physiologic measures, such as FVC, do not incorporate patient input about change and may be less than ideal when used as sole anchors in a study [13, 14, 19-21]. The overall aim of this exploratory study is to estimate the MCID values of various physiologic measures and PROMs in three different IPF cohorts using only anchor-based approach consistent with the core-criteria of the Minimally Important Difference Credibility Assessment Tool (Copyright ©2018, McMaster University) developed for evaluating anchor-based MCID studies [11]. We hypothesized that for a chronic

progressive lung disease like IPF, most patients would either be unchanged or worsened at the end of the specified follow-up period. Therefore, we calculated MCID values associated with patient worsening only.

Methods:

Data Sources

We conducted secondary analyses of data from three randomized controlled trials: Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF), AntiCoagulant Effectiveness in Idiopathic Pulmonary Fibrosis (ACE-IPF), and Prednisone, Azathioprine, and N-Acetylcysteine: A Study That Evaluates Response in Idiopathic Pulmonary Fibrosis (PANTHER-IPF) [25-27]. These three IPFnet trials were conducted by the same clinical trials group and around the same time period with similar diagnostic and adjudication process [28]. Data from these trials was obtained from the National Heart, Lung and Blood Institute (NHLBI) via the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) program. While each of these trials enrolled patients with IPF, each had different inclusion criteria and study duration: 1) the STEP-IPF trial followed patients with severe lung function impairment for 24 weeks; 2) ACE-IPF followed patients with progressive phenotype for 48 weeks; and 3) the PANTHER-IPF trial followed patients with mild to moderate impairment for 60 weeks (see Supplement Table S1 for further details). Given that the three studies had different inclusion and exclusion criteria and different follow-up time periods; three separate analyses following the same procedures were conducted for each. We used both the placebo and treatment arm patients in our analysis.

Study Measures

For our anchor, we selected the Health Transition question (SF2) in the 36-Item Short Form Survey. SF2 asks the patients to rate their health on a five point Likert scale in response to the following question: "Compared with one year ago, how would you rate your health in general now"? Possible responses to this question were as follows: (1) "much better," (2) "somewhat better," (3) "same," (4) "somewhat worse," and (5) "much worse" [29]. The SF2 is a general question that has been used in MCID determination in other studies and meets the requirements of patient reported anchor proposed by the Minimally Important Difference Credibility Assessment Tool (Copyright ©2018, McMaster University) [7, 11, 16]. Since it is not specific for a domain such as dyspnea or physical function, SF2 is a suitable anchor for all the measures of interest in the analysis. It was also available for all three studies and for all follow-up intervals. Data for other possible anchors, such as one of the PROMs or sub-domains of PROMs, was not available for all study cohorts. We analyzed patients with complete SF2 data at the end of the respective study follow-up time period.

The physiologic measures included in our analysis were FVC, total lung capacity (TLC), diffusing capacity of the lung for carbon monoxide (DLCO), six minute walk distance (6MWD). We evaluated both absolute change in percent predicted FVC and FVC in liters (L) separately. We also analyzed relative change in FVC in L which was expressed as a percentage. For DLCO we evaluated absolute difference in percent predicted DLCO and DLCO measured as

ml/min/mmHg. The STEP-IPF dataset obtained from BioLINCC did not include percent predicted values for FVC and DLCO. We used NHANES spirometry reference values to compute percent predicted values for FVC for the STEP-IPF cohort [30]. Percent predicted values for DLCO were not computed for STEP-IPF cohort. For TLC, the absolute difference in TLC in L was analyzed in ACE-IPF and PANTHER-IPF cohorts. The TLC values were not available in the STEP-IPF dataset. For 6MWD, we analyzed absolute difference in 6MWD in meters.

The PROMs we examined included Borg dyspnea scale, Short Form Health Survey 36 (SF36) physical and mental component scores, EuroQol score index and visual analogue scores, SGRQ, UCSD SOBQ and Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults (ICECAP) questionnaire. The STEP-IPF data set did not include total scores for SGRQ, SF36 physical and mental components, UCSD SOBQ, EuroQoL index and visual analogue scale or ICECAP questionnaire. We calculated the total scores for UCSD SOBQ and the EuroQol index and visual analogue scale (using the SAS code provided by EuroQol Group). We were unable to compute total scores for SGRQ, SF 36 and ICECAP in the STEP-IPF cohort due to missing components.

Statistical Analysis

All analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC) and IBM SPSS Statistics version 26 (SPSS Inc., Chicago, IL). All analyses were conducted using observed cases. If patients had missing data at follow-up, then those patients were not included in the MCID analysis. We initially performed descriptive univariate analyses for each patient measure retaining all outliers in the analysis. We calculated mean change between follow-up and baseline (score difference) of each measure for patients in each of the categories in the SF2 question.

For MCID calculation we followed a step-wise approach detailed in Supplement Appendix 1. Briefly, we assessed responsiveness of each measure with SF2 by using receiver operating curve analysis. Only those measures that met the criteria for responsiveness i.e., area under the curve (AUC) ≥ 0.70, were selected for MCID estimation. We determined the score difference of the measures from baseline to follow-up in patients who answered "somewhat worse" in response to SF2 as the MCID.

Results:

Baseline Characteristics

A total of 140 patients had follow-up data at 24 weeks in the STEP-IPF cohort, 111 patients had follow-up data at 48 weeks in the ACE-IPF cohort and 228 patients had follow-up data at 60 weeks in the PANTHER-IPF trial. Participants from all three cohorts were predominantly male (71-81%) and white (92-96%). The STEP-IPF cohort had a mean (SD) age of 68.47 (9.11) years with mean (SD) percent predicted FVC of 58.52 (15.50)% and mean (SD) DLCO of 7.92 (2.12) ml/min/mmHg (Supplement Table S2). The ACE-IPF cohort had a mean (SD) age of 66.65 (7.49) years with a mean (SD) percent predicted FVC of 61.94 (15.19)% and mean (SD) DLCO of 36.16 (12.90) % (Supplement Table S3). The PANTHER-IPF cohort had mean (SD) age of

67.05 (8.32) years with a mean (SD) percent predicted FVC of 73.81(15.05)% and DLCO of 46.18 (11.36) % (Supplement Table S4).

Response to Anchor SF2

In the STEP-IPF cohort, 110 out of the 140 patients (78.6%) were either in the "same" or in the "somewhat worse" category according to SF2 response at follow-up (Supplement Table S5). 6MWD was the only measure in the STEP-IPF cohort that met the responsiveness criteria (AUC $\geq$ 0.70) for further MCID estimation (Table 1). The AUC for other physiologic measures and PROMs in the STEP-IPF ranged from 0.55 - 0.68 (Supplement Table S5). In the ACE-IPF cohort, 98 out of the 111 patients (88.3%) with follow-up data at 48 weeks answered "same" or "somewhat worse" in response to the SF 2 question at 48 weeks (Supplement Table S6). None of the physiologic measures or the PROMs in the ACE-IPF cohort met the prespecified responsiveness criteria for further MCID determination with AUC ranging from 0.53 to 0.61 (Supplement Tables S6). In the PANTHER-IPF cohort, 175 out of 228 patients (76.8%) answered about the same or somewhat worse in response to the SF2 question (Supplement Table S7). In the PANTHER-IPF cohort, the physical component score of the SF-36 questionnaire, the total SGRQ and UCSD SOBQ scores were the only measures that met criteria for next stage of MCID calculation (Table 2). The AUC for other physiologic measures and PROMs in the PANTHER-IPF ranged from 0.47 - 0.69 (Supplement Table S7).

Anchor-based MCID Values for Worsening (Table 3)

The following measures did not meet responsiveness criteria (AUC $\geq$0.70) in any of the cohorts: FVC, TLC, DLCO, Borg dyspnea score, SF-36 mental component score, EuroQol score index and visual analogue scores, and ICECAP scores. Therefore, no MCID values were determined for these measures. 6MWD met the responsiveness criteria only in the STEP-IPF cohort, therefore, the MCID values for 6MWD was determined only at 24 weeks. SGRQ, SF-36 physical component score, and UCSD SOBQ met the responsiveness criteria only in the PANTHER-IPF cohort, therefore, MCIDs were determined only at 60 weeks interval for these measures. The mean change from baseline to follow-up (24 weeks for 6MWD and 60 weeks for the other three measures) in patients who answered "somewhat worse" in response to SF2 was selected as the MCID. MCID value for 6MWD was -74.89 meters (95% CI -93.11, -56.66) over 24 weeks. The MCID value for physical component score (PCS) of SF-36 over 60 weeks was -6.79 points (95% CI -8.66, -4.92). MCID value for total SGRQ score over 60 weeks was 10.95 points (95% CI 7.81, 14.1). MCID value for the total UCSD SOBQ score over 60 weeks was 11.38 points (95% CI 7.83, 14.93).

Discussion:

This is the first study in IPF to conduct a comprehensive exploratory analysis of multiple physiologic measures and PROMs in three different cohorts using only an anchor-based approach consistent with recently proposed standards in the MCID literature and demonstrates several key points [11]. First, the MCID estimates of 6MWD, SGRQ, SF-36, UCSD SOBQ were higher than previously calculated point estimates. These previous studies not only used different methodology, but in most instances, conducted their analyses on patients with different baseline

disease severity and with different follow-up intervals which makes direct comparison difficult. Second, in our analysis, no one measure met responsiveness criteria in more than one cohort. Third, the variable FVC, the primary end point in major trials, did not meet responsiveness criteria in any of the three cohorts. This variation in responsiveness of outcome measures may be due to random chance, different duration of follow-up compared to the anchor, study procedures, or bias; or some combination of them all. Our findings demonstrate the complexities of MCID calculation which has large implications for trial design and evaluation.

Our study's results must be understood in the context of its limitations. First, the different time periods for the three trial cohorts limited an analysis of a combined cohort and thereby restricted the sample size for the analysis. The lack of similar follow-up time also limited our ability to validate MCID values from one cohort in another cohort. Further studies in other trial cohorts using a similar anchor-based approach are needed to verify and validate the results of our study. Second, we used a single anchor, SF2, for our analysis. While general transition rating questions such as SF2 have been widely used as anchors, the results of our analysis should be confirmed with other anchors [11, 23]. Additionally, SF2 asks patient to recall their general health over the last one year which makes it prone to recall bias and using this to anchor changes over other time periods may not be ideal. Third, the anchor-based method used in our analysis is prone to regression to the mean phenomenon [10, 31]. There are no clear guidelines on which anchor-based methods to use in estimating MCID [11]. Additional consensus recommendations on the most accurate and precise anchor-based methods are warranted and would lead to further standardization of the MCID calculation. Finally, our study assessed responsiveness and estimated MCID but did not assess the validity or psychometric properties of these measures. Previous studies have evaluated convergent validity and some psychometric properties of 6MWD, SF-36, SGRQ and UCSD SOBQ in IPF [13, 14, 21, 22, 32, 33]. Even with these limitations, the MCID values estimated in our analysis represent some significant methodological strengths over prior IPF work.

We utilized a systematic approach consistent with the core criteria proposed by the recently published Minimally Important Difference Credibility Assessment Tool (Copyright ©2018, McMaster University) to evaluate MCID studies with one notable methodologic exception [11]. Most MCID studies and the afore-mentioned credibility instrument propose using a correlation coefficient (usually ≥ 0.3 or 0.5) to assess responsiveness of the change in the measure with the anchor [10, 31]. This approach is suitable for diseases such as chronic pain where patients are expected to be categorized somewhat evenly into the five-point Likert scale categories of an anchor like SF2. However, in a chronic progressive disease like IPF, most patients may fall into only two of the five anchor categories as was the case in our analysis. Therefore, using correlation coefficient may not accurately identify variables that are responsive to the anchor. Given the imbalance in categories, which was seen in all three cohorts in our study, the receiver operating curve analysis with AUC ≥ 0.70 was used to assess responsiveness of variables to a dichotomous anchor [34, 35].

Compared to previous MCID studies in IPF, we did not use distribution-based methods in our calculation. Since distribution-based methods do not take into account patient's report of their health, they essentially report the minimal detectable change (MDC). However, MDC and MCID are two different concepts as illustrated by de Vet and Terwee [24]. Previous MCID studies in

IPF have used distribution-based methods along with anchor-based methods and have reported lower point estimates when compared to our calculated values (Table 3). The MCID estimate for 6MWD at 75m in our analysis is much higher compared to previously reported values ranging from 21.7 - 45m [15, 17, 21]. The estimate for SF-36 PCS of 7 points is also higher when compared to previous values of 3 points and 5 points [14, 20]. While the difference in baseline disease severity and follow-up intervals in some of the previous studies makes direct comparison difficult, in certain cases our MCID values fall within the reported ranges of previous studies even if they are higher than the point estimates. For instance, only one study thus far has determined MCID estimates of total UCSD SOBQ scores and used the STEP-IPF cohort for their analysis [22]. They reported an MCID estimate of 8 points for both improvement and worsening with a range of 5-11 over 24 weeks using SGRQ's activity domain for anchor-based method along with distribution-based methods [22]. The UCSD SOBQ score did not meet responsiveness criteria in our analysis of STEP-IPF cohort but our reported anchor-based MCID values for UCSD SOBQ at 11.38 points over a 60 week time period using mild to moderate disease patients of the PANTHER-IPF trial is close to the reported range of 5-11 in the previous study.

Similarly, an earlier study reported an MCID of SGRQ as 7 points with a range of 5-10 using both anchor-based and distribution-based methods in IPF patients with mild to moderate severity [14]. In our analysis, we estimated higher MCID of SGRQ of 10.95 over 60-week time period for worsening using a similar mild to moderate category of patients which again within the range of the previous study but higher than the reported point estimate. However, another more recent study estimated MCID for SGRQ in IPF using mild to moderate severity patients over 52 weeks and proposed a threshold of 4-5 points for both improvement and worsening using both distribution and anchor-based methods and is much lower than our estimate [13]. Further research is needed to study the impact of MCID methodology, disease severity, follow-up interval on MCID estimation and there are efforts underway to study some of these relationships in other diseases such as Asthma [36]. A study of MCID of three questionnaires including SGRQ in COPD patients found stable MCID values over different follow up intervals ranging from 3 weeks to 12 months [37]. A large real world dataset of IPF patients, such as the newer patient registries, with patients of varying disease severity and multiple follow-up measurements at set intervals may be useful for standardized MCID research of physiologic measures and PROMs, provided they have appropriate anchors for MCID estimation [38, 39].

Conclusions:

Our study highlights the fact the anchor-based MCID estimates of 6MWD, SGRQ, SF-36, UCSD SOBQ in our study were considerably higher when compared to point estimates from previously proposed values. Further research is needed to assess MCID values of various physiologic measures and PROMs in IPF using a more current and standardized approach in different patient cohorts over different time periods to better design and evaluate clinical trials. There is further need to establish MCID of newer physiologic measures such as home spirometry and actigraphy [40, 41]. PROMs designed specifically for IPF patients are also needed to better capture the patient experience in clinical trials since PROMs like SGRQ were developed for patients with obstructive diseases. The newly proposed Living with Idiopathic Pulmonary Fibrosis (L-IPF) questionnaire is one such endeavor to better incorporate the patient experience [42]. With these

advances, future intervention trials in IPF may be better poised to accurately evaluate patient quality of life.

Acknowledgements: This Manuscript was prepared using ACE, PANTHER, STEPIPF Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the ACE, PANTHER, STEPIPF or the NHLBI. The preliminary findings of this project were presented as a poster abstract during the American Thoracic Society (ATS) International Conference in May 2020. This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Use of the Minimally Important Difference Credibility Assessment Tool, authored by Dr. Tahira Devji et al, was made under license from McMaster University, Hamilton, Canada.

References:

1. Raghu G, Richeldi L. Current approaches to the management of idiopathic pulmonary fibrosis. Respir Med. 2017;129:24-30. doi: 10.1016/j.rmed.2017.05.017.
2. Collard HR, Brown KK, Martinez FJ, Raghu G, Roberts RS, Anstrom KJ. Study design implications of death and hospitalization as end points in idiopathic pulmonary fibrosis. Chest. 2014;146(5):1256-62. doi: 10.1378/chest.14-0492.
3. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med. 2015;192(2):e3-19. doi: 10.1164/rccm.201506-1063ST.
4. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet. 2011;377(9779):1760-9. doi: 10.1016/s0140-6736(11)60405-4.
5. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071-82. doi: 10.1056/NEJMoa1402584.
6. King TE, Jr., Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2083-92. doi: 10.1056/NEJMoa1402582.
7. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15. doi: 10.1016/0197-2456(89)90005-6.
8. Jones PW, Beeh KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. Am J Respir Crit Care Med. 2014;189(3):250-5. doi: 10.1164/rccm.201310-1863PP.
9. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. Jama. 2014;312(13):1342-3. doi: 10.1001/jama.2014.13128.
10. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol. 2017;82:128-36. doi: 10.1016/j.jclinepi.2016.11.016.
11. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020;369:m1714. doi: 10.1136/bmj.m1714.
12. Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, et al. Minimally important difference estimates and methods: a protocol. BMJ Open. 2015;5(10):e007953. doi: 10.1136/bmjopen-2015-007953.
13. Swigris JJ, Wilson H, Esser D, Conoscenti CS, Stansen W, Kline Leidy N, et al. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis: insights from the INPULSIS trials. BMJ Open Respir Res. 2018;5(1):e000278. doi: 10.1136/bmjresp-2018-000278.

14. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, et al. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. Respiratory medicine. 2010;104(2):296-304. doi: 10.1016/j.rmed.2009.09.006.
15. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med. 2011;183(9):1231-7. doi: 10.1164/rccm.201007-1179OC.
16. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med. 2011;184(12):1382-9. doi: 10.1164/rccm.201105-0840OC.
17. Nathan SD, du Bois RM, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Validation of test performance characteristics and minimal clinically important difference of the 6-minute walk test in patients with idiopathic pulmonary fibrosis. Respir Med. 2015;109(7):914-22. doi: 10.1016/j.rmed.2015.04.008.
18. Nolan CM, Delogu V, Maddocks M, Patel S, Barker RE, Jones SE, et al. Validity, responsiveness and minimum clinically important difference of the incremental shuttle walk in idiopathic pulmonary fibrosis: a prospective study. Thorax. 2017. doi: 10.1136/thoraxjnl-2017-210589.
19. Prior TS, Hoyer N, Hilberg O, Shaker SB, Davidsen JR, Bendstrup E. Responsiveness and minimal clinically important difference of SGRQ-I and K-BILD in idiopathic pulmonary fibrosis. Respir Res. 2020;21(1):91. doi: 10.1186/s12931-020-01359-3.
20. Witt S, Krauss E, Barbero MAN, Muller V, Bonniaud P, Vancheri C, et al. Psychometric properties and minimal important differences of SF-36 in Idiopathic Pulmonary Fibrosis. Respir Res. 2019;20(1):47. doi: 10.1186/s12931-019-1010-5.
21. Swigris JJ, Wamboldt FS, Behr J, du Bois RM, King TE, Raghu G, et al. The 6 minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. Thorax. 2010;65(2):173-7. doi: 10.1136/thx.2009.113498.
22. Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, et al. The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. Respiratory medicine. 2012;106(10):1447-55. doi: 10.1016/j.rmed.2012.06.018.
23. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-34. doi: 10.1016/j.jclinepi.2009.08.010.
24. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. J Clin Epidemiol. 2010;63(7):804-5; author reply 6. doi: 10.1016/j.jclinepi.2009.12.015.
25. Noth I, Anstrom KJ, Calvert SB, de Andrade J, Flaherty KR, Glazer C, et al. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2012;186(1):88-95. doi: 10.1164/rccm.201202-0314OC.
26. Idiopathic Pulmonary Fibrosis Clinical Research N, Raghu G, Anstrom KJ, King TE, Jr., Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. N Engl J Med. 2012;366(21):1968-77. doi: 10.1056/NEJMoa1113354.
27. Idiopathic Pulmonary Fibrosis Clinical Research N, Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, et al. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. N Engl J Med. 2010;363(7):620-8. doi: 10.1056/NEJMoa1002110.

28. Andrade J, Schwarz M, Collard HR, Gentry-Bumpass T, Colby T, Lynch D, et al. The Idiopathic Pulmonary Fibrosis Clinical Research Network (IPFnet): diagnostic and adjudication processes. Chest. 2015;148(4):1034-42. doi: 10.1378/chest.14-2889.
29. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.
30. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. American journal of respiratory and critical care medicine. 1999;159(1):179-87.
31. Malec JF, Ketchum JM. A Standard Method for Determining the Minimal Clinically Important Difference for Rehabilitation Measures. Arch Phys Med Rehabil. 2020;101(6):1090-4. doi: 10.1016/j.apmr.2019.12.008.
32. Swigris JJ, Esser D, Wilson H, Conoscenti CS, Schmidt H, Stansen W, et al. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis. Eur Respir J. 2017;49(1). doi: 10.1183/13993003.01788-2016.
33. Swigris JJ, Streiner DL, Brown KK, Belkin A, Green KE, Wamboldt FS, et al. Assessing exertional dyspnea in patients with idiopathic pulmonary fibrosis. Respiratory medicine. 2014;108(1):181-8. doi: 10.1016/j.rmed.2013.12.009.
34. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge University Press; 2011.
35. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3.
36. Lanario J, Jones R, Hyland M, Menzies-Gow A, Mansur A, Dodd J, et al. Is the minimally clinically important difference (MCID) fit for purpose? A planned study using the SAQ. European Respiratory Journal. 2020;56(suppl 64):2241. doi: 10.1183/13993003.congress-2020.2241.
37. Alma H, de Jong C, Jelusic D, Wittmann M, Schuler M, Kocks J, et al. The minimal clinically important difference for COPD health status tools measured with global ratings of change during different time periods. European Respiratory Journal. 2016;48(suppl 60):PA873. doi: 10.1183/13993003.congress-2016.PA873.
38. Guenther A, Krauss E, Tello S, Wagner J, Paul B, Kuhn S, et al. The European IPF registry (eurIPFreg): baseline characteristics and survival of patients with idiopathic pulmonary fibrosis. Respir Res. 2018;19(1):141. doi: 10.1186/s12931-018-0845-5.
39. Wang BR, Edwards R, Freiheit EA, Ma Y, Burg C, de Andrade J, et al. The Pulmonary Fibrosis Foundation Patient Registry: Rationale, Design, and Methods. Ann Am Thorac Soc. 2020. doi: 10.1513/AnnalsATS.202001-035SD.
40. Russell AM, Adamali H, Molyneaux PL, Lukey PT, Marshall RP, Renzoni EA, et al. Daily Home Spirometry: An Effective Tool for Detecting Progression in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2016;194(8):989-97. doi: 10.1164/rccm.201511-2152OC.
41. Nathan SD, Flaherty KR, Glassberg MK, Raghu G, Swigris J, Alvarez R, et al. A Randomized, Double-Blind, Placebo-Controlled Study of Pulsed, Inhaled Nitric Oxide in Subjects at Risk of Pulmonary Hypertension Associated With Pulmonary Fibrosis. Chest. 2020;158(2):637-45. doi: 10.1016/j.chest.2020.02.016.

42. Swigris JJ, Andrae DA, Churney T, Johnson N, Scholand MB, White ES, et al. Development and Initial Validation Analyses of the Living with Idiopathic Pulmonary Fibrosis (L-IPF) Questionnaire. American Journal of Respiratory and Critical Care Medicine. 2020(ja).

Table 1. Change in measures that met responsiveness criteria over 24 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the STEP-IPF trial

Variable	SF-2 Response at 24-Week Follow-Up, Mean (SD)					AUC (95% CI)†
	Much Better (N= 7)	Somewhat Better (N=12)	About the Same (N= 46)	Somewhat worse* (N=58)	Much Worse (N=9)	
Physiologic Measures						
Δ6MWD, meters	-3.71 (42.33)	-35.08 (144.51)	-32.29 (95.46)	-74.89 (69.32)	-163.00 (122.93)	0.72 (0.61-0.83)

* The mean change in scores in the "somewhat worse" category is the MCID value for measures that meet the responsiveness criteria

† Receiver operating curve comparing dichotomous Health Transition question (SF2) response (about the "same" vs. "somewhat worse") with mean change score of variable

Δ Absolute change over 24 weeks

Abbreviations:

AUC Area under curve

6MWD Six-minute walk distance

Table 2. Change in measures that met responsiveness criteria over 60 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the PANTHER-IPF trial

Variable	SF-2 Response at 60-Week Follow-Up, Mean (SD)					AUC (95% CI)†
	Much Better (N= 11)	Somewhat Better (N=32)	About the Same (N= 101)	Somewhat worse* (N=74)	Much Worse (N=10)	
Physiologic Measures						
ΔSF36 Physical Component Score‡	8.00 (11.57)	-0.92 (6.26)	-0.06 (5.72)	-6.79 (8.07)	-5.29 (9.26)	0.75 (0.67-0.83)
Subjective Measures						
ΔTotal SGRQ Score§	-13.70 (14.55)	-0.97 (9.60)	1.35 (9.18)	10.95 (13.19)	25.86 (17.46)	0.71 (0.63-0.79)
ΔUCSD SOBQ Total Score	-4.45 (11.79)	3.19 (9.60)	0.59 (12.81)	11.38 (15.31)	34.90 (23.65)	0.72 (0.65-0.80)

* The mean change in scores in the "somewhat worse" category is the MCID value for measures that meet the responsiveness criteria

† Receiver operating curve comparing dichotomous Health Transition question (SF2) response (about the "same" vs. "somewhat worse") with mean change score of variable

‡ N missing 5 (0 much better, 1 somewhat better, 4 same, 0 somewhat worse, 0 much worse)

§ N missing 13 (0 much better, 2 somewhat better, 8 same, 3 somewhat worse, 0 much worse)

Δ Absolute change over 60 weeks

Abbreviations:

AUC Area under curve

SF36 36 Item Short Form Survey

SGRQ St. George's Respiratory Questionnaire

UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire

Table 3. Anchor-based estimates of Minimal Clinically Important Difference (MCID) for worsening in idiopathic pulmonary fibrosis from current study and comparison with MCID estimates from previous studies

Variable	Current MCID Estimation		Previous MCID Estimation‡	
	Time Period*	MCID† (95% CI)	Time Period§	MCID
6MWD, meters	24 Weeks	-74.89 (-93.11, -56.66)	48 Weeks [15]	24-45¶
			48 Weeks [17]	21.7 –37.0¶
			52 Weeks [21]	28
SF36 Physical Component Score	60 Weeks	-6.79 (-8.66, -4.92)	26 Weeks [14]	3
			No specified time period [19]	5
Total SGRQ Score	60 Weeks	10.95 (7.81, 14.1)	26 Weeks [14]	7 (5-10)#
			52 Weeks [13]	4-5
UCSD SOBQ Total Score	60 Weeks	11.38 (7.83, 14.93)	24 Weeks [22]	8 (5-11)#

* 6MWD met responsiveness criteria only in STEP-IPF cohort while the SF 36 physical component score, SGRQ and UCSD SOBQ met criteria in only the PANTHER-IPF cohort therefore time period for MCID is 24 weeks or 60 weeks

† MCID for worsening only since it was calculated as mean change in "somewhat worse" Health Transition question (SF2) group from baseline to follow-up

‡ Previous studies have estimated MCID values for worsening and improvement using both anchor-based and distribution-based methods

§ Citation in parentheses

¶ Range

Point Estimate (Range)

Abbreviations:

6MWD Six-minute walk distance

SF36 36 Item Short Form Survey

SGRQ St. George's Respiratory Questionnaire

UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire

Supplement

An Updated Approach to Determine Minimal Clinically Important Differences In Idiopathic Pulmonary Fibrosis.

Mohleen Kang, MD; Srihari Veeraraghavan, MD; Greg S Martin, MD, MSc; Jordan A Kempker, MD, MSc

Content

Appendix 1: Detailed approach for MCID determination

Appendix 2: Tables

S1. Description of the three randomized control trials of idiopathic pulmonary fibrosis in adult patients used for secondary data analysis
S2. Baseline characteristics of STEP-IPF patients with Health Transition question (SF2) data at 24 Weeks
S3. Baseline characteristics of ACE-IPF patients with Health Transition question (SF2) data at 48 weeks
S4. Baseline characteristics of PANTHER IPF patients with Health Transition question (SF2) data at 60 weeks
S5. Change in physiologic and patient reported outcome measures over 24 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the STEP-IPF trial
S6. Change in physiologic and patient reported outcome measures over 48 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the ACE-IPF trial
S7. Change in physiologic and patient reported outcome measures over 60 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the PANTHER-IPF trial.

Appendix 1: Detailed Approach for MCID Determination

The following analysis plan will be followed for analyzing each of the three datasets (ACE-IPF, PANTHER-IPF and STEP-IPF) separately

Selected Anchor: Health Transition question (SF2) in SF-36 questionnaire

> "Compared with one year ago, how would you rate your health in general now"?
>
> (1) "much better,"
>
> (2) "somewhat better,"
>
> **(3) "same,"**
>
> **(4) "somewhat worse," and**
>
> (5) "much worse."
>
> The difference between the third and the fourth responses (i.e., "same" and "somewhat worse") was considered to represent minimal clinically important difference for worsening.

Follow-up interval:

ACE-IPF	48 weeks 111 patients
PANTHER-IPF	60 weeks 228 patients
STEP-IPF	24 weeks 140 patients

General Approach:

We conducted receiver operating curve analysis to assess responsiveness of the change in variable of interest with the anchor SF2 (dichotomous variable "same" and "somewhat worse"). Variables with both an area under the curve (AUC) ≥ 0.70 and with appropriate direction of response i.e., worsening scores with worsening response to SF2 question were selected for further MCID determination. We calculated mean change (score difference between follow-up and baseline) of each variable for patients in each of the categories in the SF2 question. The score difference of the variable from baseline to follow-up in patients who answered "somewhat worse" in response to SF2 was selected as MCID.

Steps:

1. Calculate score change in variable of interest (VOI) by subtracting follow-up scores from baseline score.

2. Assess SF2 response at follow-up time period and compute mean score change in VOI within each SF2 response category
3. Compute area under the curve (AUC) between SF2 ("same" and "somewhat worse" categories only) and VOI score change
    a. value≥ 0.70 selected to next step.
4. Calculate anchor-based MCID Values: Mean Score Change in "somewhat worse" SF2 Group from Baseline to Follow-up

Appendix 2:

Table S1. Description of the three randomized control trials of idiopathic pulmonary fibrosis in adult patients used for secondary data analysis

Study Name	Enrollment	Number of Centers	Sample Size	Disease Severity/ Phenotype	Treatment	Time Period	Findings
STEP-IPF	2007-2009	14 US centers	180	Severe (DLCO <35%)	Sildenafil	Two 12 weeks: 1st placebo vs. sildenafil, 2nd sildenafil both groups	No significant improvement in 6MWD between groups
ACE-IPF	2009-2011	22 US centers	145	Progressive phenotype (either worsening dyspnea or absolute decline of FVC ≥10%, DLCO decline ≥15%, arterial oxygen saturation decline ≥5% or worsening radiographic findings)	Warfarin	48 weeks	Study terminated earlier since patients on warfarin had higher mortality, hospitalization and severe side effects
PANTHER-IPF	2009-2011	25 US centers	155	Mild to moderate (FVC ≥50% and DLCO ≥30%)	Prednisone,azathioprine and n-acetylcysteine	60 weeks	Trial stopped early. Treatment group with increased mortality, hospitalizations and adverse events

Abbreviations:
STEP-IPF Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis
DLCO Diffusing capacity for carbon monoxide
6MWD Six minute walk distance
FVC Forced vital capacity
ACE-IPF AntiCoagulant Effectiveness in Idiopathic Pulmonary Fibrosis
PANTHER-IPF Prednisone, Azathioprine, and N-Acetylcysteine: A Study That Evaluates Response in Idiopathic Pulmonary Fibrosis

Table S2. Baseline characteristics of STEP-IPF patients with Health Transition question (SF2) data at 24 Weeks

Characteristic	**Total Subjects (N=140)**	**Treatment (N=69)**	**Placebo (N=71)**
Demographic Characteristics			
Age, years	68.47 (9.11)	69.66 (8.51)	67.35 (9.56)
Male, N (%)	114 (81.43)	58 (84.06)	56 (78.87)
White, N (%) †	130 (92.86)	63 (91.3%)	67 (94.36)
Hispanic or Latino, N (%)	10 (7.14)	6 (8.70)	4 (5.63)
Clinical Characteristics			
Past or Current Smoker, N (%)	112 (80.00)	54 (78.26)	58 (81.69)
Pulmonary Function Testing			
FEV1 %*	65.10 (16.23)	63.95 (16.88)	66.18 (15.65)
FEV1, L	1.95 (0.57)	1.91 (0.57)	1.98 (0.57)
FVC %*	58.52 (15.50)	56.37 (14.97)	60.51 (15.82)
FVC, L	2.37 (0.76)	2.29 (0.72)	2.45 (0.79)
TLC, L	3.68 (1.09)	3.62 (1.00)	3.74 (1.17)
DLCO, ml/min/mmHg	7.92 (2.12)	7.77 (1.94)	8.06 (2.28)
6MWD, meters	280.07 (112.27)	269.67 (99.88)	290.47 (123.28)
Quality of Life Scores			
Pre Borg Dyspnea Scale	0.79 (1.17)	0.84 (1.09)	0.74 (1.26)
Post Borg Dyspnea Scale	4.03 (8.09)	3.51 (1.68)	4.54 (11.25)
Total UCSD Score‡	46.82 (20.79)	49.61 (21.61)	44.21 (19.81)
EuroQol Visual Analog Score	69.34 (16.61)	68.41 (16.53)	70.26 (16.76)
Euroqol Index Score	0.80 (0.13)	0.79 (0.15)	0.80 (0.11)

*All values are percent predicted according to age, height, sex and race/ethnicity

† N Missing 10: 6 in treatment group and 4 in placebo

‡ N missing 22: 12 in treatment arm and 10 in placebo group

All other variables with N missing <5

FEV1 Forced expiratory volume in 1 second

FVC Forced vital capacity
TLC Total lung capacity
DLCO Diffusing capacity for carbon monoxide
6MWD Six-minute walk distance
SF36 36 Item Short Form Survey
SGRQ St. George's Respiratory Questionnaire
UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire
ICECAP Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults

Table S3. Baseline characteristics of ACE-IPF patients with Health Transition question (SF2) data at 48 weeks

Characteristic	Total Subjects (N=111)	Treatment (N=54)	Placebo (N=57)
Demographic Characteristics			
Age in years, mean (SD)	66.65 (7.49)	66.65 (7.35)	66.64 (7.68)
Male, N (%)	79 (71.17)	35 (64.81)	44 (77.19)
White, N (%)	103 (92.79)	50 (92.59)	53 (92.98)
Minority, N (%)	14 (21.61)	7 (12.96)	7 (12.28)
Clinical Characteristics			
Past or Current Smoker, N (%)	83 (74.77)	37 (68.52)	46 (80.70)
Prednisone Treatment at Randomization, N (%)	27 (24.32)	10 (18.52)	17 (29.82)
Years since IPF diagnosis, mean (SD)	0.24 (0.43)	0.19 (0.39)	0.30 (0.46)
Pulmonary Function Testing, mean (SD)			
FVC, L	2.48 (0.78)	2.42 (0.77)	2.53 (0.79)
FVC %*	61.94 (15.19)	61.77 (15.94)	62.10 (14.58)
FEV1, L	2.03 (0.61)	2.00 (0.61)	2.06 (0.61)
FEV1 %*	65.68 (15.53)	66.28 (16.86)	65.12 (14.28)
DLCO, ml/min/mmHg	10.77 (4.41)	10.69 (4.03)	10.84 (4.77)
DLCO %*	36.16 (12.90)	36.39 (12.08)	35.94 (13.75)
TLC, L	3.75 (0.99)	3.71 (0.93)	3.78 (1.05)
TLC%*	59.38 (58.72)	59.40 (13.22)	59.36 (13.60)
6MWD, meters	297.90 (128.20)	303.73 (118.35)	292.59 (137.40)
Quality of Life Scores, mean (SD)			

Pre Borg Dyspnea Scale	0.59 (0.91)	0.48 (0.92)	0.69 (0.89)
Post Borg Dyspnea Scale	2.54 (1.54)	2.28 (1.53)	2.78 (1.53)
Total SGRQ Score	45.13 (16.19)	41.99 (16.08)	48.10 (15.85)
SF36 Physical Component Score	38.13 (8.66)	40.55 (8.37)	35.84 (8.36)
SF 36 Mental Component Score	53.37 (8.33)	53.51 (7.37)	53.24 (9.21)
Total UCSD SOBQ Score	34.30 (20.71)	29.35 (17.61)	38.98 (22.42)
EuroQoL Index Score	0.76 (0.19)	0.78 (0.19)	0.75 (0.19)
EuroQol Visual Analog Score	74.90 (15.40)	75.72 (15.16)	74.14 (15.72)
ICECAP Score	0.86 (0.09)	0.87 (0.09)	0.85 (0.10)

*All values are percent predicted according to age, height, sex and race/ethnicity
FEV1 Forced expiratory volume in 1 second
FVC Forced vital capacity
TLC Total lung capacity
DLCO Diffusing capacity for carbon monoxide
6MWD Six-minute walk distance
SF36 36 Item Short Form Survey
SGRQ St. George's Respiratory Questionnaire
UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire
ICECAP Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults

Table S4. Baseline characteristics of PANTHER IPF patients with Health Transition question (SF2) data at 60 weeks

Characteristic	Total Subjects (N=228)	Treatment (N=110)	Placebo (N=118)
Demographic Characteristics			
Age, years	67.05 (8.32)	67.50 (8.43)	66.63 (8.24)
Male, N (%)	173 (75.88)	87 (79.09)	86 (72.88)
White, N (%)	219 (96.05)	105 (95.45)	114 (96.61)
Minority, N (%)	15 (6.58)	7 (6.36)	8 (6.78)
Clinical Characteristics			
Past or Current Smoker, N (%)	164 (72.25)	78 (71.56)	86 (72.88)
Years since IPF diagnosis, median (IQR)	0.72 (0.29-1.54)	0.62 (0.29-1.54)	0.79 (0.28 -1.81)
Pulmonary Function Testing			
FVC, L	2.96 (0.80)	2.99 (0.82)	2.94 (0.78)
FVC % *	73.81 (15.05)	73.65 (15.70)	73.95 (14.49)
FEV1, L	2.44 (0.64)	2.44 (0.64)	2.44 (0.64)
FEV1 % *	78.85 (16.29)	78.06 (16.42)	79.59 (16.21)
DLCO, ml/min/mmHg	13.63 (3.76)	13.58 (3.69)	13.67 (3.85)
DLCO %*	46.18 (11.36)	45.70 (10.74)	46.63 (11.94)
TLC, L	4.38 (1.03)	4.45 (1.05)	4.33 (1.01)
6MWD, meters	383.71 (107.47)	385.34 (111.03)	382.21 (104.54)
Quality of Life Scores			
Pre Borg Dyspnea Scale	0.41 (0.81)	0.44 (0.88)	0.38 (0.74)
Post Borg Dyspnea Scale	2.41 (1.66)	2.27 (1.36)	2.53 (1.89)
Total SGRQ Score	38.21 (16.80)	38.72 (16.41)	37.74 (17.21)

SF36 Physical Component Score	41.18 (9.15)	41.47 (8.95)	40.91 (9.36)
SF 36 Mental Component Score	54.65 (7.90)	53.87 (8.19)	55.38 (7.58)
Total UCSD Score	25.91 (17.54)	25.35 (16.80)	26.42 (18.26)
EuroQoL Index Score	0.83 (0.16)	0.82 (0.17)	0.84 (0.14)
EuroQol Visual Analog Score	78.25 (14.51)	78.38 (15.30)	78.13 (13.76)
ICECAP Score	0.88 (0.08)	0.88 (0.08)	0.88 (0.09)

*All values are percent predicted according to age, height, sex and race/ethnicity
FEV1 Forced expiratory volume in 1 second
FVC Forced vital capacity
TLC Total lung capacity
DLCO Diffusing capacity for carbon monoxide
6MWD Six-minute walk distance
SF36 36 Item Short Form Survey
SGRQ St. George's Respiratory Questionnaire
UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire
ICECAP Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults

Table S5. Change in physiologic and patient reported outcome measures over 24 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the STEP-IPF trial

Variable	All Patients, Mean (SD) (N=140)	SF2 Response at 24-Week Follow-Up, Mean (SD)					AUC (95% CI)†
		Much Better (N= 7)	Somewhat Better (N=13)	About the Same (N= 48)	Somewhat worse (N=62)	Much Worse (N=10)	
Physiologic Measures							
ΔFVC%*	-2.72 (4.66)	-1.89 (4.49)	0.55(3.91)	-1.78 (4.52)	-3.68 (3.96)	-7.46 (6.75)	0.62 (0.51-0.72)
ΔFVC, L	-0.11 (0.20)	-0.09 (0.19)	0.04 (0.18)	-0.07 (0.19)	-0.15 (0.16)	-0.33 (0.25)	0.62 (0.51-0.72)
RFVC, %	-4.88 (8.38)	-3.44 (8.49)	1.13 (8.00)	-2.58 (7.50)	-6.98 (7.35)	-13.03 (10.67)	0.65 (0.55-0.76)
ΔDLCO, ml/min/mmHg	-0.48 (1.53)	-0.47 (1.54)	0.00 (1.30)	-0.16 (1.42)	-0.66 (0.51)	-1.65 (2.10)	0.55 (0.44-0.66)
Δ6MWD, meters	**-58.66 (96.44)**	**-3.71 (42.33)**	**-35.08 (144.51)**	**-32.29 (95.46)**	**-74.89 (69.32)**	**-163.00 (122.93)**	**0.72 (0.61-0.83)**
Subjective Measures							
ΔPre Borg Dyspnea Score	1.62 (11.65)	0.07 (0.93)	0.04 (0.78)	0.13 (1.01)	3.50 (17.27)	0.06 (3.17)	0.58 (0.47-0.69)
ΔPost Borg Dyspnea Score	1.06 (14.27)	-14.29 (35.62)	-1.00 (1.73)	4.10 (19.23)	0.76 (1.89)	2.00 (3.20)	0.59 (0.49-0.70)
ΔEuroQol Index Score	-0.05 (0.17)	0.09 (0.21)	-0.05 (0.20)	0.01 (0.12)	-0.08 (0.16)	-0.20 (0.21)	0.68 (0.58-0.78)
ΔEuroQol Visual Analogue Score	-1.32 (19.69)	26.57 (24.11)	8.62 (17.57)	1.52 (16.97)	-5.84 (16.57)	-21.89 (0.42)	0.66 (0.55-0.76)
ΔUCSD SOBQ Total Score‡	5.92 (16.6)	-8.00 (7.94)	-5.33 (13.66)	2.60 (9.92)	7.90 (19.10)	24.00 (13.83)	0.55 (0.43-0.68)

*All values are percent predicted according to age, height, sex and race/ethnicity

† Receiver operating curve comparing dichotomous SF2 response (about the "same" vs. "somewhat worse") and mean change score of variable
‡ N missing 39 (4 "much better", 7 "somewhat better", 13 "same", 12 "somewhat worse", 3 "much worse"). For all other variables the SF2 columns had N missing ≤ 5.
Δ Absolute change over 24 weeks
Abbreviations:
AUC Area under curve
FVC Forced vital capacity
RFVC Relative difference in forced vital capacity
TLC Total lung capacity
DLCO Diffusing capacity for carbon monoxide
6MWD Six-minute walk distance
UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire

Table S6. Change in physiologic and patient reported outcome measures over 48 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the ACE-IPF trial

Variable	All Patients, Mean (SD), (N=111)	SF2 Response at 48-Week Follow-Up, Mean (SD)				AUC (95%CI)†
		Somewhat Better (N=6)	About the Same (N= 57)	Somewhat worse (N=41)	Much Worse (N=7)	
Physiologic Measures						
ΔFVC%*	-2.11 (6.61)	-1.00 (7.18)	-1.23 (5.61)	-1.73 (6.14)	-12.39 (8.94)	0.54 (0.41-0.66)
ΔFVC, L	-0.09 (0.26)	-0.04 (0.29)	-0.05 (0.20)	-0.08 (0.25)	-0.50 (0.40)	0.55 (0.42-0.69)
RFVC, %	-3.38 (10.61)	-0.21 (10.15)	-1.66 (8.33)	-3.23 (10.78)	-20.95 (12.53)	0.54 (0.42-0.66)
ΔTLC, L	0.00 (0.59)	0.15 (0.74)	0.06 (0.55)	-0.05 (0.60)	-0.34 (0.69)	0.58 (0.47-0.70)
ΔDLCO%*‡	-4.34 (9.80)	-0.23 (5.02)	-2.15 (6.38)	-7.22 (13.54)	-12.29 (7.13)	0.60 (0.48-0.73)
ΔDLCO, ml/min/mmHg‡	-1.34 (2.94)	-0.23 (1.32)	-0.64 (1.82)	-2.23 (4.11)	-3.83 (2.31)	0.61 (0.48-0.74)
Δ6MWD, meters	-37.50 (114.96)	-18.81 (48.96)	-19.60 (119.23)	-49.31 (107.52)	-129.20 (119.90)	0.60 (0.48-0.71)
Subjective Measures						
ΔPre Borg Dyspnea Score	0.25 (1.29)	-0.30 (1.64)	0.01 (1.18)	0.57 (1.36)	1.10 (0.89)	0.60 (0.48-0.72)
ΔPost Borg Dyspnea Score§	0.49 (1.82)	-0.40 (1.34)	0.23 (1.60)	0.51 (1.55)	4.00 (2.65)	0.56 (0.44-0.69)
ΔSF36 Physical Component Score	-2.49 (7.15)	2.80 (9.21)	-1.60 (6.38)	-2.76 (6.73)	-12.77 (4.86)	0.55 (0.43-0.67)
ΔSF36 Mental Component Score	-1.56 (7.33)	1.81 (5.45)	-0.16 (6.65)	-2.81 (8.04)	-8.70 (4.35)	0.57 (0.45-0.68)
ΔEuroQol	-0.03 (0.18)	0.07 (0.08)	-0.01 (0.18)	-0.05 (0.20)	-0.07 (0.21)	0.54 (0.42-0.66)

Index Score						
ΔEuroQol Visual Analogue Score	-5.86 (15.19)	1.80 (11.01)	-3.21 (11.80)	-7.46 (17.35)	-23.57 (17.49)	0.59 (0.47-0.70)
ΔTotal SGRQ Score	4.54 (11.62)	0.57 (21.17)	2.41 (8.44)	5.20 (11.16)	20.73 (14.27)	0.58 (0.46-0.70)
ΔUCSD SOBQ Total Score	8.42 (16.65)	8.83 (12.25)	5.54 (10.86)	6.98 (16.65)	40.00(27.45)	0.53 (0.41-0.64)
ΔICECAP Score	-0.02 (0.10)	0.04 (0.12)	0.00(0.06)	-0.02 (0.11)	-0.16 (0.08)	0.59 (0.47-0.71)

*All values are percent predicted according to age, height, sex and race/ethnicity

† Receiver operating curve comparing dichotomous SF2 response (about the "same" vs. "somewhat worse") and mean change score of variable

‡ N missing 12 (0 "somewhat better", 3 "same", 8 "somewhat worse", 1 "much worse")

§ N missing 15 (1 "somewhat better", 6 "same", 6 "somewhat worse", 2 "much worse")

For all other variables the SF2 columns had N missing ≤ 5.

Δ Absolute change over 48 weeks

Abbreviations:

AUC Area under curve

FVC Forced vital capacity

RFVC Relative difference in forced vital capacity

TLC Total lung capacity

DLCO Diffusing capacity for carbon monoxide

6MWD Six-minute walk distance

SF36 36 Item Short Form Survey

SGRQ St. George's Respiratory Questionnaire

UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire

ICECAP Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults

Table S7. Change in physiologic and patient reported outcome measures over 60 weeks by Health Transition question (SF2) categorical responses in patients with idiopathic pulmonary fibrosis in the PANTHER-IPF trial.

Variable	All Patients, Mean (SD) (N=228)	SF2 Response at 60-Week Follow-Up, Mean (SD)					AUC (95% CI)†
		Much Better (N= 11)	Somewhat Better (N=32)	About the Same (N= 101)	Somewhat worse (N=74)	Much Worse (N=10)	
Physiologic Measures							
ΔFVC%*	-4.09 (6.75)	1.18 (7.31)	-2.64 (6.77)	-2.73 (5.80)	-6.65 (5.97)	-10.09 (11.13)	0.68 (0.60-0.76)
ΔFVC, L	-0.16 (0.28)	0.05 (0.29)	-0.10 (0.29)	-0.10 (0.23)	-0.27 (0.25)	-0.40 (0.44)	0.68 (0.60-0.76)
RFVC, %	-5.74 (9.99)	1.93 (10.81)	-3.46 (9.63)	-3.68 (8.06)	-9.46 (8.91)	-15.82 (19.43)	0.69 (0.61-0.77)
ΔTLC, L	-0.15 (0.49)	-0.01 (0.36)	0.07 (0.51)	-0.12 (0.44)	-0.27 (0.49)	-0.56 (0.63)	0.64 (0.55-0.73)
ΔDLCO%*	-4.28 (7.62)	0.38 (6.17)	-4.86 (5.91)	-3.32 (7.18)	-5.16 (8.06)	-14.78 (10.07)	0.59 (0.50-0.68)
ΔDLCO, ml/min/mmHg	-1.27 (2.21)	0.23 (1.72)	-1.37 (1.65)	-0.98 (1.99)	-1.58 (2.44)	-4.46 (2.91)	0.58 (0.50-0.67)
Δ6MWD, meters	-34.53 (100.53)	-2.32 (46.60)	-18.23 (119.36)	-22.64 (77.60)	-52.11 (113.71)	-229.00 (91.69)	0.57 (0.48-0.66)
Subjective Measures							
ΔPre Borg Dyspnea Score	0.29 (1.39)	-0.27 (0.52)	-0.39 (0.90)	0.39 (1.06)	0.51 (1.87)	1.30 (1.72)	0.51 (0.42-0.60)
ΔPost Borg Dyspnea Score	0.32 (0.92)	-1.27 (2.04)	-0.47 (1.78)	0.26 (1.61)	0.83 (2.01)	3.63 (1.60)	0.61 (0.52-0.69)
ΔSF36 Physical Component Score	**-2.25 (8.07)**	**8.00 (11.57)**	**-0.92 (6.26)**	**-0.06 (5.72)**	**-6.79 (8.07)**	**-5.29 (9.26)**	**0.75 (0.67-0.83)**
ΔSF36 Mental	-1.35 (8.61)	-0.29 (12.78)	-1.43 (9.66)	-0.41 (6.05)	-0.56 (7.61)	-17.35 (13.32)	0.47 (0.38-0.56)

Component Score							
ΔEuroQol Index Score	-0.04 (0.17)	-0.01 (0.19)	-0.06 (0.13)	0.01 (0.17)	-0.07 (0.14)	-0.29 (0.32)	0.66 (0.58-0.75)
ΔEuroQol Visual Analogue Score‡	-1.50 (17.19)	15.90 (16.39)	1.07 (10.75)	1.53 (16.35)	-5.79 (16.14)	-25.67 (18.86)	0.63 (0.54-0.72)
ΔTotal SGRQ Score§	**4.54 (13.74)**	**-13.70 (14.55)**	**-0.97 (9.60)**	**1.35 (9.18)**	**10.95 (13.19)**	**25.86 (17.46)**	**0.71 (0.63-0.79)**
ΔUCSD SOBQ Total Score	**5.72 (15.94)**	**-4.45 (11.79)**	**3.19 (9.60)**	**0.59 (12.81)**	**11.38 (15.31)**	**34.90 (23.65)**	**0.72 (0.65-0.80)**
ΔICECAP Score¶	-0.01 (0.09)	0.04 (0.08)	-0.01 (0.08)	0.01 (0.07)	-0.03 (0.09)	-0.16 (0.19)	0.69 (0.61-0.78)

*All values are percent predicted according to age, height, sex and race/ethnicity

† Receiver operating curve comparing dichotomous SF2 response (about the "same" vs. "somewhat worse") and mean change score of variable

‡ N missing 19 (1 "much better", 5 "somewhat better", 9 "same", 3 "somewhat worse", 1 "much worse")

§ N missing 13 (0 "much better", 2 "somewhat better", 8 "same", 3 "somewhat worse", 0 "much worse")

¶ N missing 12 (0 "much better", 0 "somewhat better", 6 "same", 6 "somewhat worse", 0 "much worse")

For all other variables the SF2 columns had N missing ≤ 5.

Δ Absolute change over 60 weeks

Abbreviations:

AUC Area under curve

FVC Forced vital capacity

RFVC Relative difference in forced vital capacity

TLC Total lung capacity

DLCO Diffusing capacity for carbon monoxide

6MWD Six-minute walk distance

SF36 36 Item Short Form Survey

SGRQ St. George's Respiratory Questionnaire

UCSD SOBQ The University of California, San Diego Shortness of Breath Questionnaire
ICECAP Investigating Choice Experiences for the Preferences of Older People Capability Instruments for Adults